

Computer Science Department

TECHNICAL REPORT

CONSTRAINED TOTAL LEAST SQUARES PROBLEMS
AND THE SMALLEST PERTURBATION OF A SUB-
MATRIX WHICH LOWERS THE RANK

By

James Demmel
September 1985

Technical Report #174

NEW YORK UNIVERSITY



Department of Computer Science
Courant Institute of Mathematical Sciences
251 MERCER STREET, NEW YORK, N.Y. 10012

WIS COMPSCI TR-174
Demmel, James Weldon
Constrained total
least squares
c.2



CONSTRAINED TOTAL LEAST SQUARES PROBLEMS
AND THE SMALLEST PERTURBATION OF A SUB-
MATRIX WHICH LOWERS THE RANK

By

James Demmel
September 1985

Technical Report #174

NEW YORK UNIVERSITY
COURANT INSTITUTE-LIBRARY
251 Mercer St. New York, N.Y. 10012

**Constrained Total Least Squares Problems
and the Smallest Perturbation of a Submatrix which Lowers the Rank**

James Weldon Demmel
Courant Institute of Mathematical Sciences
251 Mercer Str.
New York, NY 10012

Abstract

Let $T = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ be a partitioned rectangular matrix. We exhibit perturbations of D of smallest norm which lower the rank of T . We apply this result to generalize the total least squares problem $Sx=b$ when only a rectangular subset of the data $[S,b]$ is considered uncertain.

1. Introduction

Let $T = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ be a partitioned rectangular matrix. We exhibit perturbations of D of smallest Frobenius or two norm which cause T to have a lower rank. This generalizes a recent result of Stewart [2] who found the smallest rank lowering perturbation of D when $T=[C|D]$. It may be, of course, that the rank of T is independent of D , as in the 2 by 2 example

$$T = \begin{bmatrix} 0 & 1 \\ 1 & D \end{bmatrix}$$

which is of full rank for all scalars D . In such cases our method will report that no perturbation of D exists. In such cases a small change in the matrix can cause a large change in the answer. If

$$T(\epsilon) = \begin{bmatrix} -\epsilon & 1 \\ 1 & D \end{bmatrix}$$

then D must equal ϵ^{-1} to lower the rank. As ϵ decreases to 0, ϵ^{-1} grows to infinity, until at $\epsilon=0$ no rank lowering perturbation of D exists. Thus the problem of determining the smallest rank lowering perturbation of D can be quite ill-conditioned. This is in marked contrast to

the classical problem of finding a smallest rank lowering perturbation to the matrix as a whole. In this case the answer is the smallest singular value of the matrix, and it can change no more than the matrix itself changes (measured in the 2-norm).

We apply this result to generalize the procedure of total least squares (TLS) as described in [1]. TLS arises when trying to solve a least squares problem $Sx=b$ where S has more rows than columns. In traditional least squares, one considers the entries of S to be exact and the entries of b to be contaminated with noise, and one interprets the residual norm

$$\min_x \|Sx-b\| = \min_{b+r \in R(S)} \|r\| ,$$

(where $R(S)$ is the span of the columns of S and $\|\cdot\|$ is the Euclidean vector norm) as a measure of this contamination. In TLS, one considers the entries of both S and b to be contaminated with noise, and interprets the value of

$$\min_{b+r \in R(S+E)} \|[E,r]\|_F$$

(where $\|\cdot\|_F$ is the Frobenius norm) as a measure of that contamination. If the noise in some components of S and b are to be weighted differently, one can compute

$$\min_{b+r \in R(S+E)} \|W_1[E,r]W_2\|_F$$

where W_1 and W_2 are nonsingular diagonal weight matrices. Golub and Van Loan show that the value of this minimum is given by $\sigma_{\min}(W_1[S,b]W_2)$, the smallest singular value of $W_1[S,b]W_2$.

We generalize this TLS procedure to the case where only some rows and columns of S and b are contaminated. If, for example, only the last m rows of S and b and the last n columns of S are contaminated, applying our result to

$$T = W_1[S,b]W_2 = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where D is m by $n+1$ provides a measure of the contamination in these entries. Another interpretation of this result is as a solution of the constrained total least squares problem, since by considering A and B to be exact, they provide constraints on the solution vector.

2. Main Result

There are a number of cases in our result depending of the ranks of submatrices obtained in a sequence of transformations performed on T . Therefore, we present our result as an algorithm for determining the smallest rank lowering perturbation of D ; comments proving the correctness will be inserted in the algorithm.

Our algorithm will make frequent use of the singular value decomposition (SVD), which we will write as follows: $A = U\Sigma V^T$, with U and V orthogonal and $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_k)$, $\sigma_1 \geq \dots \geq \sigma_k \geq 0$. One can read the rank of A directly from its SVD: it is the number of nonzero σ_i . In terms of the SVD, it is well known that a smallest perturbation of A which lowers its rank to r is $-U\text{diag}(0, \dots, 0, \sigma_{r+1}, \dots, \sigma_k)V^T$, where r must be less than or equal to k . The Frobenius norm of this perturbation is $(\sum_{i=r+1}^k \sigma_i^2)^{1/2}$ and its two norm is σ_{r+1} . Our algorithm will also make use of row compressions and column compressions, which can be computed in several ways. Finding a orthogonal matrix U which column compresses a matrix C to the left (resp. to the right) means finding U such that CU has $\text{rank}(C)$ linearly independent columns as its leftmost (resp. rightmost) columns, and the other columns zero. It can be computed either by doing a QR decomposition of C^T or by a singular value decomposition of C [1]. Similarly, row compressing C upwards (resp. downwards) means finding a orthogonal U such that UC has $\text{rank}(C)$ linearly independent rows at the top (resp. bottom) and the other rows zero.

Our approach depends on a result of Stewart's which we state here:

Algorithm 1: [2] Given $T = [C|D]$, C an n by m_1 matrix and D an n by m_2 matrix, compute a smallest perturbation of D which lowers the rank of T to $r < n$ (if one exists).

1) Find U to row compress C upwards. Let $c = \text{rank}(C)$ be the number of nonzero rows of UC . Write

$$UT = \begin{bmatrix} C_1 & D_1 \\ 0 & D_2 \end{bmatrix}$$

where D_1 is c by m_2 and D_2 is $n-c$ by m_2 .

2) If $c > r$, the rank of T is greater than r independent of D , since the first c rows of UC are independent. Terminate the algorithm.

3) (At this point $c \leq r$.) Clearly, $\text{rank}(T) = c + \text{rank}(D_2)$, so the problem reduces to finding a smallest perturbation of D_2 which lowers its rank to $r-c$. This problem is solved by the SVD. Let $U_D \Sigma_D V_D^T$ be the SVD of D_2 , and let $d_2 = \text{rank}(D_2)$. If $c + d_2 \leq r$, the rank of T is already less than or equal to r . Otherwise the smallest perturbation of D that lowers the rank of T to r has Frobenius norm $(\sum_{l=r-c+1}^{d_2} \sigma_l^2)^{1/2}$, and equals $U^T \begin{bmatrix} 0 \\ \delta D_2 \end{bmatrix}$, where δD_2 is the smallest perturbation of D_2 which lowers the rank of D_2 to $r-c$. Terminate the algorithm.

Note that the rank of T must lie between c and $c + \min(n-c, m_2)$ for all D .

Now consider the more general case where

$$T = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A is m_1 by n_1 and D is m_2 by n_2 . We assume $m_1 \leq n_1$; other consider T^T . To illustrate how we might reduce this general case to the case covered by Algorithm 1, let us consider the special case where A is of full rank m_1 , and in particular let us assume that the first \tilde{m}_1 by m_1 submatrix of A is nonsingular. Accordingly partition

$$T = \begin{bmatrix} A_1 & A_2 & B \\ C_1 & C_2 & D \end{bmatrix}$$

where A_1 is m_1 by m_1 and nonsingular. Now premultiply T by the nonsingular matrix

$$\begin{bmatrix} I & 0 \\ -C_1 A_1^{-1} & I \end{bmatrix}$$

to obtain the matrix of equal rank

$$\begin{bmatrix} A_1 & A_2 & B \\ 0 & C_2 - C_1 A_1^{-1} A_2 & D - C_1 A_1^{-1} B \end{bmatrix}.$$

At this point, since A_1 has full rank m_1 , it is clear that

$$\text{rank}(T) = m_1 + \text{rank}[C_2 - C_1 A_1^{-1} A_2 \mid D - C_1 A_1^{-1} B],$$

so that the smallest perturbation of D that lowers the rank of T to r is the smallest perturbation of D that lowers the rank of $[C_2 - C_1 A_1^{-1} A_2 \mid D - C_1 A_1^{-1} B]$ to $r - m_1$ (if $r \geq m_1$; otherwise no such perturbation of D exists). This subproblem is solved by algorithm 1. Computing $C_1 A_1^{-1} B$ can be quite ill-conditioned if A_1 is nearly singular; as illustrated by the example in the introduction, it is unavoidable.

The following algorithm generalizes this approach to the case where A is not of full rank:

Algorithm 2: Given $T = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ where A is m_1 by n_1 and D is m_2 by n_2 , compute a smallest perturbation of D which lowers the rank of T to $r < \min(m_1 + m_2, n_1 + n_2)$ (if one exists).

1) Find an orthogonal U_1 which row compresses A upwards and an orthogonal U_2 which column compresses the resulting rows leftward. In other words $U_1 A U_2$ has a nonsingular matrix A_1 in its first $a = \text{rank}(A)$ rows and columns, and the other rows and columns are zero. Then transform and partition T as follows:

$$T_1 = \begin{bmatrix} U_1 & 0 \\ 0 & I_{m_2} \end{bmatrix} \cdot T \cdot \begin{bmatrix} U_2 & 0 \\ 0 & I_{n_2} \end{bmatrix} = \begin{bmatrix} A_1 & 0 & B_1 \\ 0 & 0 & B_2 \\ C_1 & C_2 & D \end{bmatrix}$$

where the first block row $[A_1 \mid 0 \mid B_1]$ and first block column $[A_1^T \mid 0 \mid C_1^T]^T$ may be null if $a=0$, where the second block row $[0 \mid 0 \mid B_2]$ may be null if $a=m_1$, and where the second block column $[0 \mid 0 \mid C_2^T]^T$ may be null if $a=n_1$.

2) If $a=m_1$ so that the second block row is null, we have reduced the problem to the special case discussed above. Premultiply T_1 by

$$\begin{bmatrix} I_{m_1} & 0 \\ -C_1 A_1^{-1} & I_{m_2} \end{bmatrix}$$

to get

$$T_2 = \begin{bmatrix} A_1 & 0 & B_1 \\ 0 & C_2 & D - C_1 A_1^{-1} B_1 \end{bmatrix}$$

so the rank of T is $\text{rank}(T_2) = a + \text{rank}[C_2 | D - C_1 A_1^{-1} B_1]$. Let $c = \text{rank}(C_2)$. Applying Theorem 1, we see that if $r < a + c$, no perturbation of D exists which lowers the rank to r . Otherwise, apply Theorem 1 to compute the smallest perturbation of D that makes $\text{rank}[C_2 | D - C_1 A_1^{-1} B_1] = r - a$. Terminate the algorithm.

3) Now $a < m_1$ so that B_2 in T_1 is nonnull. Find an orthogonal U_3 which row compresses B_2 upwards and an orthogonal U_4 which column compresses the resulting rows leftward. In other words $U_3 B_2 U_4$ has a nonsingular matrix B_{21} in its first $b = \text{rank}(B_2)$ rows and columns, and the other rows and columns are zero. Then transform and partition T_1 as follows:

$$T_3 = \begin{bmatrix} I_a & 0 & 0 \\ 0 & U_3 & 0 \\ 0 & 0 & I_{m_2} \end{bmatrix} \cdot T_1 \cdot \begin{bmatrix} I_{n_1} & 0 \\ 0 & U_4 \end{bmatrix} = \begin{bmatrix} A_1 & 0 & B_{11} & B_{12} \\ 0 & 0 & B_{21} & 0 \\ 0 & 0 & 0 & 0 \\ C_1 & C_2 & D_1 & D_2 \end{bmatrix}.$$

Since the third block row is zero it does not contribute to the rank so we drop it. Also, we swap the first and second block rows and move the third block column to the first column to get the matrix of equal rank

$$T_4 = \begin{bmatrix} B_{21} & 0 & 0 & 0 \\ B_{11} & A_1 & 0 & B_{12} \\ D_1 & C_1 & C_2 & D_2 \end{bmatrix}.$$

Let $c = \text{rank}(C_2)$. Note that in T_4 the first block row and block column may be null if $b=0$, and the second block row and block column may be null if $a=0$, and the last block column may be null if $b=n_2$.

4) If $b=n_2$, so that the last block column of T_4 is null, then $\text{rank}(T)$ is independent of D and we are done. This is because T_4 has rank

$$\text{rank} \left(\begin{bmatrix} I_b & 0 & 0 \\ 0 & I_a & 0 \\ -D_1 B_{21}^{-1} & 0 & I_{m_2} \end{bmatrix} \cdot T_4 \right) = \text{rank} \left(\begin{bmatrix} B_{21} & 0 & 0 \\ B_{11} & A_1 & 0 \\ 0 & C_1 & C_2 \end{bmatrix} \right) = a + b + c$$

which is independent of D . Terminate the algorithm.

5) Now $b < n_2$ so D_2 is nonnull. If $a+b=0$, so that the first two block rows and block columns of T_4 are null, we have

$$T_4 = [C_2 | D_2]$$

which is the case covered by Theorem 1. Then $\text{rank}(T)$ is at least $c = \text{rank}(C_2)$ for any D , and if $r \geq c$ Theorem 1 provides a perturbation δD_2 of D_2 of smallest norm which reduces the rank of T_4 to r . A smallest perturbation of D is in turn given by $[0 | \delta D_2] U_4^T$. Terminate the algorithm.

6) Now $b < n_2$ and $a+b > 0$. This is the special case illustrated earlier, since the $a+b$ by $a+b$ matrix

$$\begin{bmatrix} B_{21} & 0 \\ B_{11} & A_1 \end{bmatrix}$$

is nonsingular. Therefore we premultiply T_4 by

$$\begin{bmatrix} I_{a+b} & 0 \\ -[D_1 | C_1] \begin{bmatrix} B_{21} & 0 \\ B_{11} & A_1 \end{bmatrix}^{-1} & I_{m_2} \end{bmatrix}$$

getting

$$T_5 = \begin{bmatrix} B_{21} & 0 & 0 & 0 \\ B_{11} & A_1 & 0 & B_{12} \\ 0 & 0 & C_2 & D_2 - C_1 A_1^{-1} B_{12} \end{bmatrix}.$$

so that the rank of T is

$$\text{rank}(T) = \text{rank}(T_5) = a + b + \text{rank}([C_2 | D_2 - C_1 A_1^{-1} B_{12}]).$$

Applying Theorem 1 we see that $\text{rank}(T) \geq a+b+c$ for any D , so our problem has a solution only if $a+b+c \leq r$. Proceeding as in Algorithm 1, Let U_5 row compress C_2 upwards and partition

$$U_5[C_2|D_2 - C_1A_1^{-1}B_{12}] = \begin{bmatrix} C_{21} & D_{11}-X_1 \\ 0 & D_{22}-X_2 \end{bmatrix}$$

where C_{21} has full rank c . Let δD_{22} be a smallest perturbation of D_{22} that lowers the rank of $D_{22}-X_2$ to $r-a-b-c$. Then a smallest norm lowering perturbation of D is given by

$$[0|U_5^T \begin{bmatrix} 0 \\ \delta D_{22} \end{bmatrix}]U_4^T.$$

Terminate the algorithm.

Note that we have also shown that

$$\text{rank}(T) \geq a+b+c = \text{rank}(A) + \text{rank}(B_2) + \text{rank}(C_2)$$

independent of D , where we interpret any of a , b or c as zero if the corresponding submatrix is null. In addition, a , b and c may be determined by only using unitary transformations on T , whereas determining the smallest rank lowering perturbation of D requires nonunitary (and possibly quite ill-conditioned) transformations.

3. References

- [1] G. Golub, C. Van Loan, *Matrix Computations*, Johns Hopkins Press, Baltimore, 1983
- [2] G. W. Stewart, *A generalization of the Eckart-Young approximation theorem*, University of Maryland, Computer Science Technical Report, TR-1307, Sept. 1983

Constrained total
least squares

NYU COMPSCI TR-174
Demmel, James Weldon
Constrained total
least squares

OCT 15 1986

A fine will be charged for each day the book is kept overtime.

GAYLORD 142
PRINTED IN U.S.A.

